Bayesian Network Fundamentals

Semantics and Factorization

Simple Example

Let's first define a student example. In this example we have five random variables(RV), which noted with upper case of initial letter.

- Grade
- Course Difficulty
- Intelligence
- SAT grade
- Reference Letter

Every RV has two possible values except Grade.

Then we can construct a graph as follow, it can be regard as a model of dependency.
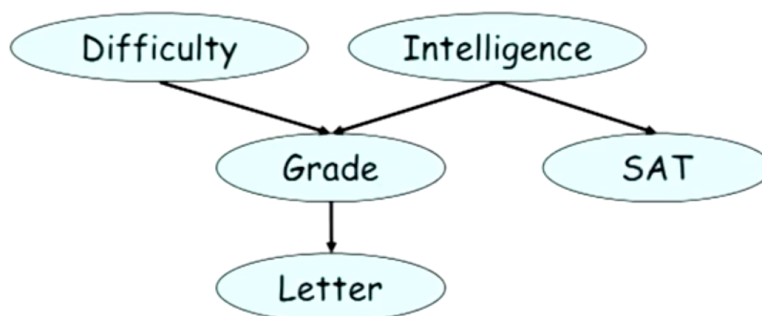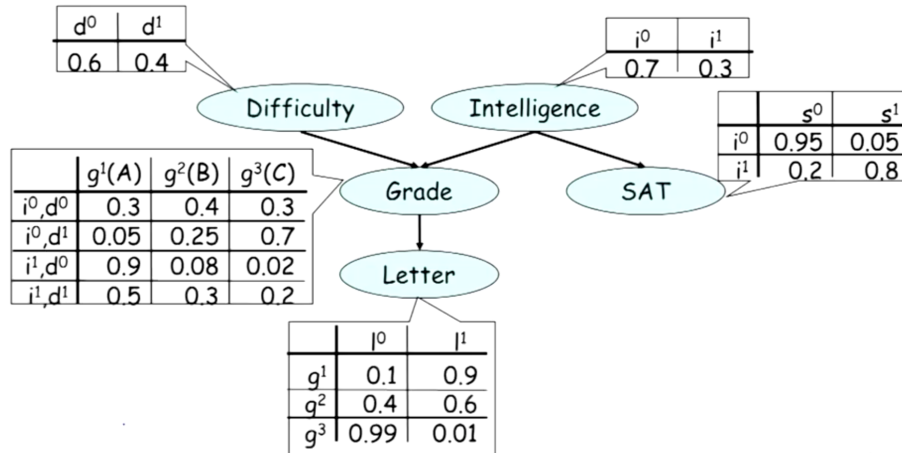

Figure 1: graph

Also you can come up with a graph with different construct. It's dependence on how you understand and represent the world.

We can use CPD(conditioned prob. distribution) to notate every node in this graph to turn this graph to probability distribution. Following picture shows a example of CPD assignment of graph.

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	$g^1(A)$	$g^2(B)$	$g^3(C)$
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Chain Rule for Bayesian Networks

As shown in picture, joint distribution can be factorized as products of other factors

Chain Rule for Bayesian Networks
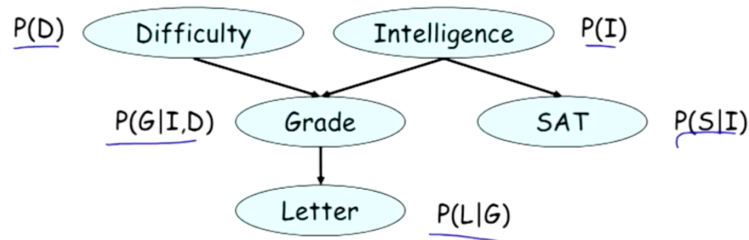
$$P(D,I,G,S,L) = P(D)\, P(I)\, P(G|I,D)\, P(S|I)\, P(L|G)$$

Bayesian Network

- A Bayesian network is:
    - A DAG G whose nodes represent the RVs $X_1, ..., X_n$
    - For each node X_i a CPD $P(X_i | Par_G(X_i))$
- BN represents a joint distribution via the chain rule for Bayesian networks
  $P(X_1...X_n) = \prod_i P(X_i | Par_G(X_i))$

P Factorizes over G

- Let G be a graph over $Z_1...X_n$
- P factorizes over G if $P(XX_1..X_N) = \prod_i P(X_i|Par_G(X_i))$.

Reasoning Patterns

Take previous student example again

Causal Reasoning

Now we have $P(l^1) = 0.5$. If we observed that $I = 0$. Then we have $P(l^1|i^0) = 0.39$.

It is not hard to explain. If a student not good in intelligence, then it will lower the prob to get higher grade, which will lead to decrease of letter.

This pattern reasoning called causal reasoning, intelligence is the cause of letter.

Or we can say if we observed ancestor, the prob of child will be changed.

Evidential Reasoning

Now considered the case we observed the grade. Originally we have $P(d^1) = 0.4$ and then we observed $G = 3$. Then $P(d^1|g^3) = 0.63$.

This pattern called evidential reasoning. If we observed child of a RV, the CPD of this RV might be changed.

Intercausal Reasoning

Now thinking that we have $P(i^1) = 0.3$. Then we know the grade is B, $G = 2$, the prob $P(i^1|g^2) = 0.175$. After that we also know the $D = 1$, the condition prob distribution becomes $P(i^1|g^2, d^1)$.

From above example, we know that when two RVs have same result. These two RVs will affect each other given their result.

Flow of Probabilistic Influence

When can X influence Y, given no observations:

- $X \rightarrow Y$. **YES**.
- $X \leftarrow Y$. **YES**.
- $X \rightarrow W \rightarrow Y$. **YES**.
- $X \leftarrow W \leftarrow Y$. **YES**.
- $X \leftarrow W \rightarrow Y$. **YES**.
- $X \rightarrow W \leftarrow Y$. **NO**.

The last structure called **v-structure**.

Active Trail: A trail $X_1 - ... - X_n$ is active if it has no v-structures $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$

Can X influence Y, given evidence about Z:

	$W \in Z$	$W \notin Z$
$X \rightarrow Y$	YES	YES
$X \leftarrow Y$	YES	YES
$X \rightarrow W \rightarrow Y$	YES	NO
$X \leftarrow W \leftarrow Y$	YES	NO
$X \leftarrow W \rightarrow Y$	YES	NO
$X \rightarrow W \leftarrow Y$	NO(if W and its all descendant not in Z) YES(else)	YES

example: D can affect S if G or L observed

Active Trails Active Trail: A trail $X_1 - ... - X_n$ is active given Z if:

- For any v-structure $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ we have that X_i or one of its descendants $\in Z$
- No other X_i is in Z